Understanding Difference in Differences Method

A Deep Dive into Econometric Techniques

Biswajit Palit, Sneha Roy

Universitat Bonn

September 19, 2024

Table of Contents

Table of Contents

Difference in Differences Estimator

- Difference in difference method involves identifying the effect of a distinct intervention or treatment, often the implementation of a new law.
- The approach compares the change in mean outcomes before and after the treatment for groups affected by the intervention with the corresponding difference for unaffected groups.

Difference in Differences Estimator

The baseline model of DiD for a 2 x 2 case is as follows :

$$Y_{igt} = \gamma_0 + \gamma_1.\alpha_g + \gamma_2.\delta_t + \beta T_{gt} + \epsilon_{igt}$$

- $T_{gt} = \alpha_g \ . \ \delta_t$
- where $\delta_t = 1$ in the post-period and $\delta_t = 0$ in the pre-period
- $\alpha_g = 1$ if treated and $\alpha_g = 0$ if not treated
- $\alpha_g \ . \ \delta_g = 1$ if treated and in the post-period and $= 0$ otherwise.
- ϵ_{igt} is the error.

Difference in Difference Assuptions

Assumptions of DiD Model :

- DiD relies on the assumption that, in absence of treatment, the unobserved differences in the outcome variable between treatment and control groups are the same overtime. This assumption is called the **parallel trends assumption**.
- The normal OLS assumptions must hold.
- DiD does not rely on assumption of **homogeneous treatment effects** When treatment effects are homogeneous, DD estimation yields average treatment effect on the treated (ATT). i.e $\hat{\beta} = \mathrm{E}[\beta]$
- When there are heterogenous treatment effect over time (within treated units), DD estimate of treatment effect may depend on choice of evaluation window.

Difference in Differences Estimator

$$\bar{Y}^{\textbf{Control}}_{\textbf{Post}} - \bar{Y}^{\textbf{Control}}_{\textbf{Pre}} = \gamma_2$$

$$\bar{Y}^{\textbf{Treatment}}_{\textbf{Post}} - \bar{Y}^{\textbf{Treatment}}_{\textbf{Pre}} = \gamma_2 + \beta$$

So now, we subtract the time effect from the difference within treated states between pre and post treatment period, which isolates out the treatment effect.

$$(\bar{Y}^{\text{Treatment}}_{\text{Post}} - \bar{Y}^{\text{Treatment}}_{\text{Pre}}) - (\bar{Y}^{\text{Control}}_{\text{Post}} - \bar{Y}^{\text{Control}}_{\text{Pre}}) = \hat{\beta}$$

	Treatment	Control	Diff
Pre-Treatment	$\gamma_0 + \gamma_1$	γ_0	γ_1
Post-Treatment	$\gamma_0 + \gamma_1 + \gamma_2 + \beta$	$\gamma_0 + \gamma_2$	$\gamma_1 + \beta$
Diff	$\gamma_2 + \beta$	γ_2	$\hat{\beta}$

Intuitively, diff-in-diff estimation is just a comparison of 4 cell-level means

Only one cell is treated: Treatment$\times$Post-Program

Difference in Differences Estimation

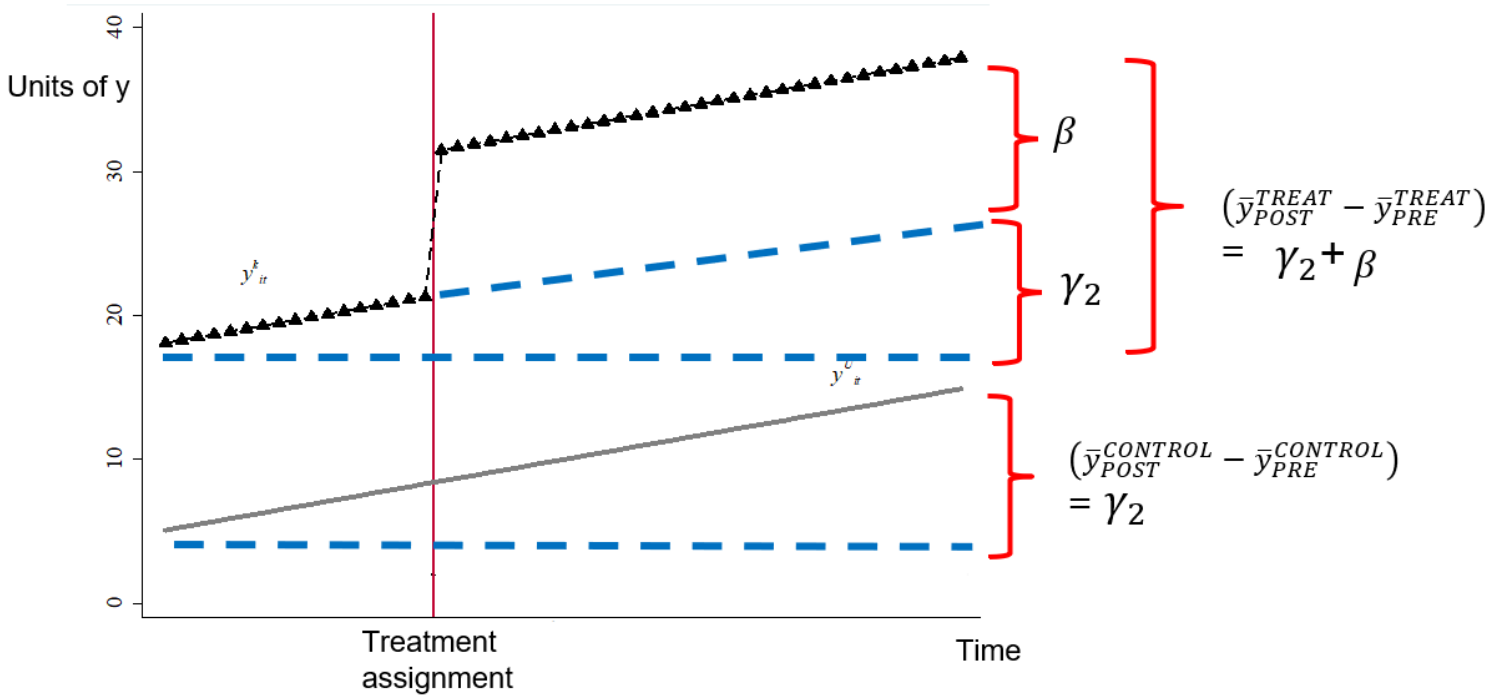

Figure 1: ATT Estimation using DiD

Table of Contents

Structure of the Data

- We extract micro data on women from CPS(MORG) for the years 1980 to 2000. We focus on all women between the ages 25 and 50.
- We extract information on weekly earnings, education, age, and state of residence.
- We define wage as log of weekly earnings. , We get ($50 * 21 = 1050$) state-year cells.
- We use age and education dummies (Up to Grade 10, High School, Masters' Degree, Doctorate Degree) as covariates at the individual level data.

CPS Data Aggregation

Step 1 - Covariate Adjustment:

- In the first step, we run a regression using micro-data where the outcome variable (y_{igt}) is regressed on individual-level control variables (w_{igt}).
- Now we get residuals of each observation and take the mean of residuals at every state-time cell to get $\bar{Y}_{gt}$.

Step 2 - Treatment Effect Estimation:

- In the second step, we use the estimated covariate-adjusted group-time means ($\bar{Y}_{gt}$) obtained in the first step. The key equation here is Equation 1:

$$\bar{Y}_{gt} = \alpha_g + \delta_t + \beta \cdot T_{gt} + \epsilon_{gt} \tag{1}$$

CPS Data Simulations: Placebo Treatment Assignment

Agenda

Our research introduces staggered placebo laws to different select states (the treatment group) for different years between 1985 to 1995.

Procedure

- In the first step, we choose at random exactly half the states (25) and designate them as treatment states.
- Then for each designated treated state, we pick a year from a uniform distribution between 1985 to 1995 and designate them as the treatment starting year for that state.
- The intervention variable T_{gt} is then defined as a dummy variable which equals 1 for all the aggregated state-time cells, in a treated state after the treatment date, 0 otherwise.

Ensuring Parallel Trends by Visualization

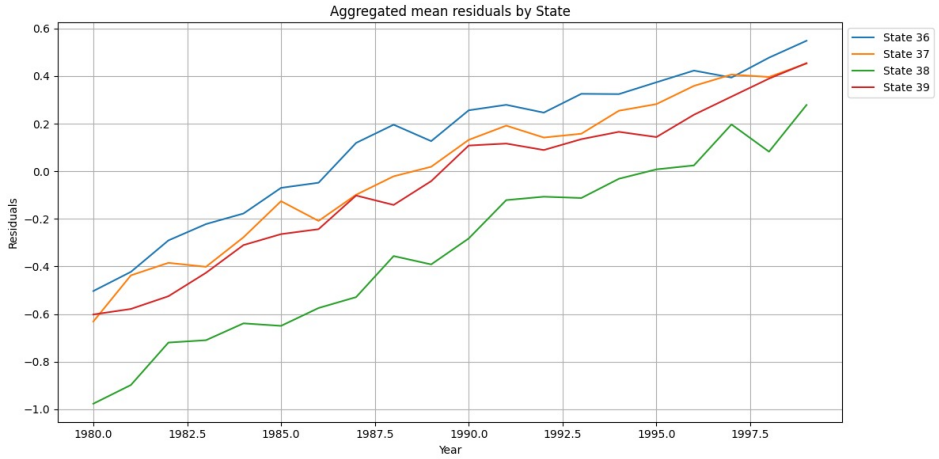

Figure 2: Trend of the Outcome Variable

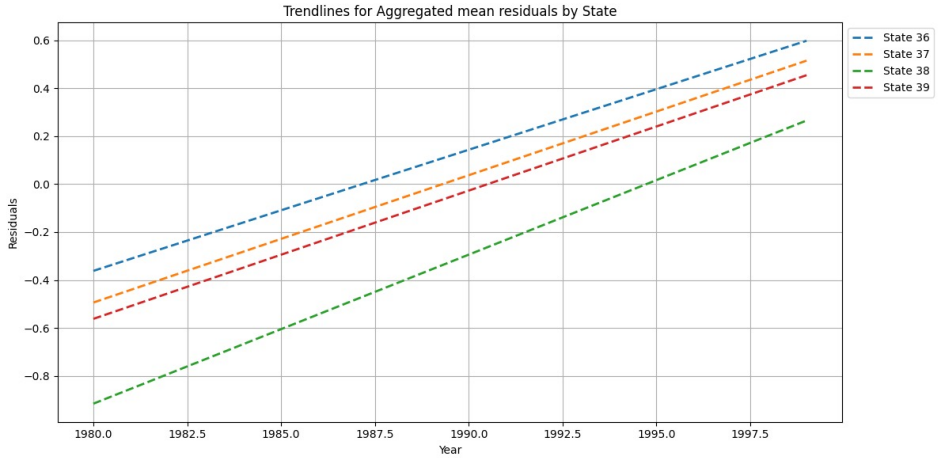

Figure 3: Ensuring Parallel Trend

Ensuring Parallel Trends by Construction

To ensure that the treated and control groups follow parallel trends, we have **randomized the assignment of placebo treatments to states**.

Randomization helps us to achieve the causal relationships by ensuring that the assignment of the treatments is purely based on chance and any differences in the outcomes of the 2 groups is by the virtue of the treatment only and no other systematic bias.

CPS Data Simulations: Placebo Treatment Assignment

Estimation

We can then estimate equation (1) using the traditional OLS method on these placebo laws. To understand how well conventional DiD performs, we can repeat this Monte Carlo simulation for 1000 times, each time drawing new laws at random.

Expectation

Since there is no law which is under effect, the traditional OLS estimation should reject the null hypothesis of $\beta = 0$,roughly 5 percent of the time. If OLS were to provide consistent standard errors, we would expect to reject the null hypothesis of no effect ($H_0 = 0$) roughly 5 percent of the time when we use a threshold of 1.96 for the absolute t-statistic.

Results Table

OLS Summary Table :

Data	CPS Aggregate
Method	OLS (assuming iid)
Rejection Rate ($\beta = 0$)	0.297 (0.014)
Avg Bias	-0.0003
Avg RMSE	0.009
Avg Std Err	0.009

Table 1: Sample Statistics of β distribution for 1000 simulations of nominal 5 percent size test

Simulation standard errors are shown in parentheses. The standard error of an estimated rejection rate is $se(\hat{r}) = \sqrt{(\hat{r})(1-\hat{r})/num\_simulations - 1}$.

Wage Correlogram

We estimate the first, second and third autocorrelation coefficients for the mean state year residuals of wages on state and year dummies.

The autocorrelation coefficients are obtained by a simple OLS regression of the residuals on the corresponding lagged residuals, therefore imposing common autocorrelation parameters for all states.

	Coefficient	Std Error	t	P value$> \|t\|$
Residual lag 1	0.34***	0.033	10.465	0.000
Residual lag 2	0.0008	0.034	0.023	0.982
Residual lag 3	0.1064***	0.032	3.284	0.001

Table 2: Coefficients of regression on the residual lags

Wage Correlogram

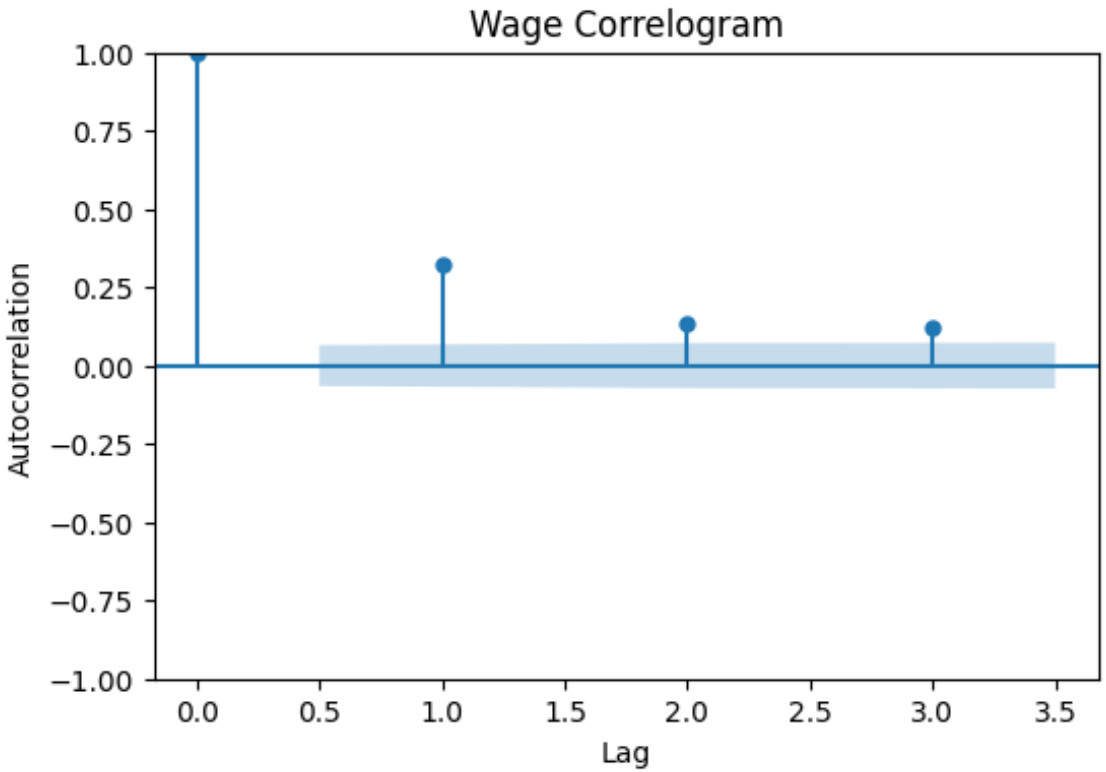

Figure 4: ACF Plot of the Wage Correlogram

Table of Contents

Data Generating Process (DGP)

We employ a specific model known as the AR(1) model, which represents our true underlying population process.
The AR(1) model is defined as follows:

$$y_{gt} = \alpha_g + \delta_t + \rho \cdot y_{gt-1} + e_{gt}$$

Where:

- y_{gt}: Observed outcome at state g and time t.
- α_g and δ_t: State and time effects, respectively, drawn from a $\mathcal{N}(0, 1)$ distribution.
- ρ: Autoregressive coefficient.
- y_{gt-1}: Lagged outcome.
- e_{gt}: Randomly chosen errors for each observation from $\mathcal{N}(0, 1)$.

Monte Carlo Result Table

Data	Method	Rej. Rate	Avg Std Err	Bias	RMSE
AR(1) ($\rho = 0$)	OLS (iid)	0.053 (0.007)	0.13	0.002	0.14
AR(1) ($\rho = 0.2$)	OLS (iid)	0.103 (0.009)	0.135	0.003	0.16
AR(1) ($\rho = 0.4$)	OLS (iid)	0.175 (0.01)	0.142	0.003	0.19
AR(1) ($\rho = 0.5$)	OLS (iid)	0.225 (0.013)	0.149	0.003	0.21
AR(1) ($\rho = 0.6$)	OLS (iid)	0.275 (0.014)	0.161	0.002	0.25
AR(1) ($\rho = 0.8$)	OLS (iid)	0.432 (0.015)	0.233	-0.001	0.49

Table 3: Rejection rates for tests of nominal 5 percent size with placebo treatments in the generated AR(1) data

Monte Carlo Results Visualization

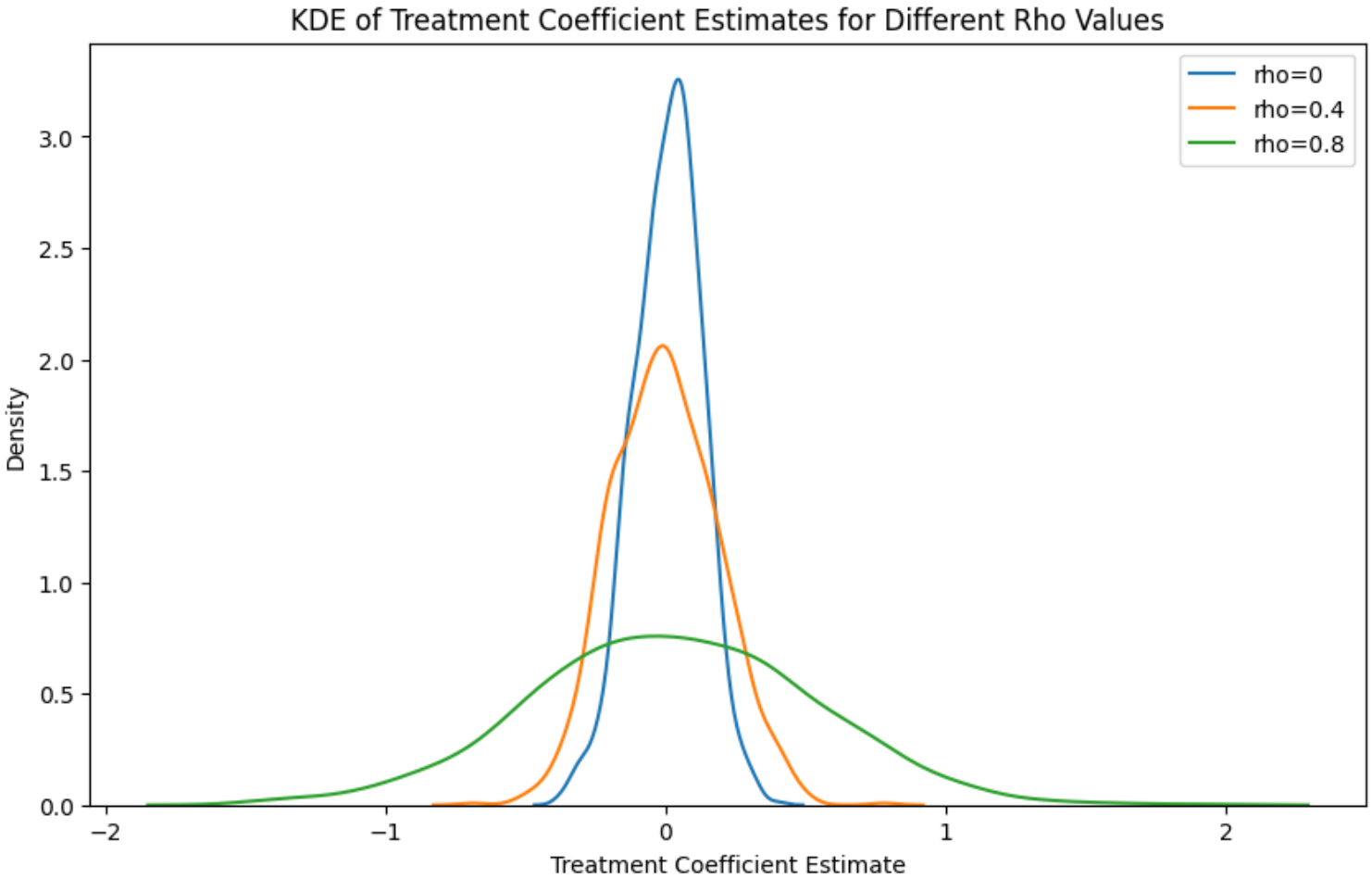

Figure 5: β distribution for varying rhos

Table of Contents

Cluster Robust Standard Errors

This method of estimating the variance-covariance matrix is consistent in the presence of any correlation pattern within states over time, even under the assumption of cross sectional heteroskedaticity.
The formula for the cluster-robust variance matrix ($\hat{\sigma}^2_{CR}$) is given as:

$$\hat{\sigma}^2_{CR} = (X'X)^{-1} \left(\sum_{g=1}^{G} X_g u_g u_g' X_g' \right) (X'X)^{-1}$$

Where:

- X: Matrix of independent variables (year dummies, state dummies and treatment dummy)
- X_g: Regressor matrix for group g.
- u_g: Vector of regression residuals for group g.

Cluster Robust Standard Errors

The traditional CRSE does not include for small g bias correction and the estimator assumes $\mathcal{N}(0, 1)$ distribution to calculate the critical values. With a small number of G, the critical values should be calculated from a t distribution with G-1 degrees of freedom.

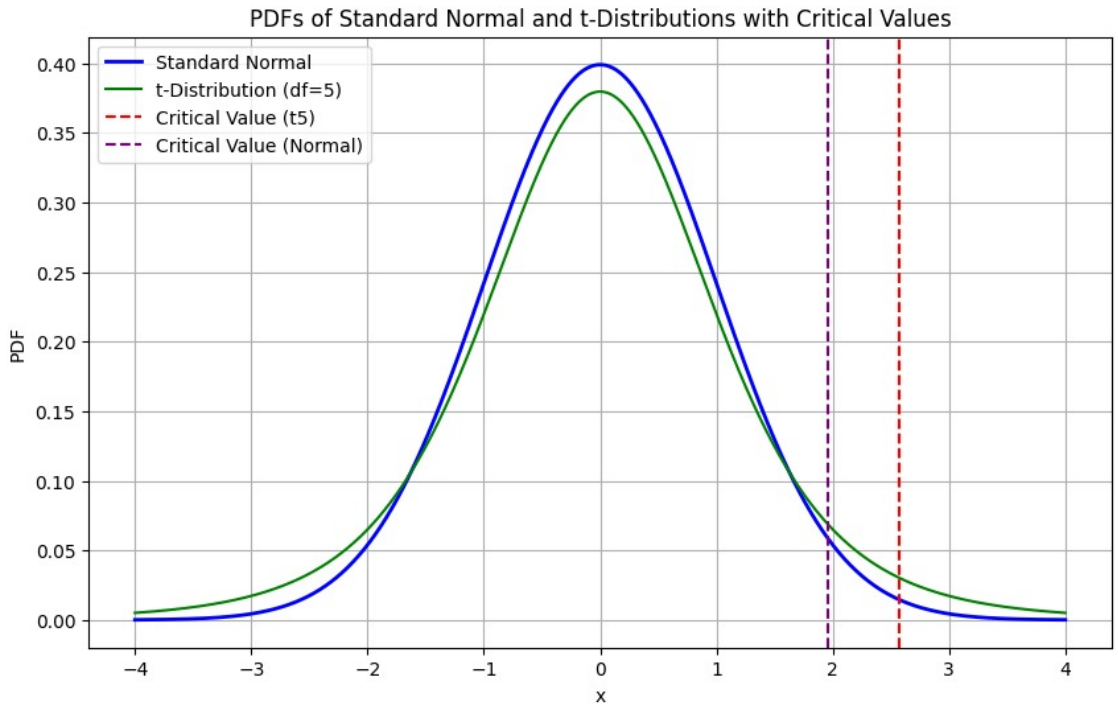

Figure 6: Standard Normal vs T(5) distribution

CRSE Simulation Result Table

Data	Method	Rej.Rate	Avg Std Err	Bias	RMSE
CPS Agg.	CRSE,N(0,1)	0.049 (0.006)	0.017	0.0001	0.016
CPS Agg.	BC-CRSE,t(G-1)	0.042 (0.006)	0.017	0.0001	0.016
AR(1) ($\rho = 0.8$)	CRSE,N(0,1)	0.044 (0.006)	0.342	0.001	0.327
AR(1) ($\rho = 0.8$)	BC-CRSE,t(G-1)	0.048 (0.006)	0.34	0.003	0.338

Table 4: Rejection rates for tests of nominal 5 percent size using CRSE methods

Wild Cluster Bootstrapping

- This approach involves empirically estimating the distribution of the test statistic through bootstrap resampling.
- Clusters of residuals are resampled, representing groups of data points.
- The unique aspect is the scaling of these resampled residuals using random constants drawn from a 2-point auxiliary distribution of -1 and 1 with mean 0 and unit variance.

Wild Cluster Bootstrapping Simulation Result Table

Data	Method	Rej. Rate	Avg Std Err	Bias	RMSE
CPS Agg	wild cluster bootstrapping	0.046 (0.007)	0.017	0.003	0.018
AR(1) ($\rho = 0.4$)	wild cluster bootstrapping	0.047 (0.006)	0.17	-0.003	0.16
AR(1) ($\rho = 0.8$)	wild cluster bootstrapping	0.055 (0.007)	0.34	-0.001	0.317

Table 5: Rejection rates for tests of nominal 5 percent size using wild cluster bootstrapping

Feasible Generalized Least Squares (FGLS)

In this approach, we assume an autoregressive (AR) process, particularly a homogenous AR(2) error structure across states, which accounts for the correlation in the shocks.
The FGLS procedure involves several key steps:

- We estimate the model using Ordinary Least Squares (OLS), as done previously.
- We estimate the parameters for the lagged AR(2) process using the OLS residuals assuming a homogenous AR(2) error structure across states.
- With these AR(2) parameter estimates, we apply standard Generalized Least Squares (GLS) linear transformations to the variables in the model:

$$y^*_{gt} = \hat{Y}_{gt} - \rho_1 \hat{Y}_{g,t-1} - \rho_2 \hat{Y}_{g,t-2}$$
$$X^*_{gt} = X_{gt} - \rho_1 X_{g,t-1} - \rho_2 X_{g,t-2}$$

Feasible Generalized Least Squares (FGLS)

- Finally, we re-estimate the model, including the transformed variables, using OLS on the adjusted data.

The GLS transformation involves adjusting the data by applying linear transformations based on the estimated AR(2) parameters. This transformation helps to decorrelate the serially correlated errors and ensures that the model assumptions are met.

FGLS Bias Correction

Intuition

- As iterated by Hansen (2007), when the number of time periods is small, the conventional estimators fail to capture the true behaviour of the true residuals.
- The alteration of behaviour of the residuals , disrupts the correlation pattern of the data, leading to biased estimates of ρ

Correction

- After estimating ρ_1 and ρ_2 by OLS, an iterative loop is generated where in each iteration a new set of ρ_1' and ρ_2' is generated based on the autocovariance matrix of the current estimated rhos.
- The parameters are updated by the difference between the plim estimate of the ρ_1' , ρ_2', and the current estimates which is the bias.
- If the bias is below a specified tolerance level, the loop ends, otherwise the loop goes on until it reaches max iterations, then it goes back to the initial OLS estimates.

FGLS Simulation Result Table

Data	Method	Rej.Rate	Avg Std Err	Bias	RMSE
CPS Agg.	FGLS	0.041 (0.006)	0.017	0.001	0.016
CPS Agg.	BC-FGLS	0.043 (0.006)	0.017	-0.001	0.016
AR(1) ($\rho = 0.8$)	FGLS	0.07 (0.008)	0.2	-0.0003	0.2
AR(1) ($\rho = 0.8$)	BC-FGLS	0.05 (0.007)	0.2	-0.002	0.2

Table 7: Rejection rates for tests of nominal 5 percent size using FGLS methods

Table of Contents

Power to Detect Real Effect

- Our findings in the last section iterate that controlling test size is not the major concern in DiD designs. But the power of these processes to detect real effects is a major concern.
- In this section, we have introduced a 0.02 log point, i.e, we have added 0.02 to Y_{gt} only for the treated state time cells. The idea is to simulate a 2 percent effect to the data and then we have proceeded with the data aggregation as before.
- We want to calculate the power of the power of the test by calculating the number of times the null hypothesis of no effect ($H_0 : \beta = 0$) gets rejected against the alternate hypothesis of ($H_1 : \beta = 0.02$) out of 500 simulations.

Power Analysis of Different Methods

Data	Method	Power	Avg Std Err	Bias
CPS Agg.	OLS, iid	0.7 (0.02)	0.009	0.001
CPS Agg.	CRSE, N(0,1)	0.184 (0.01)	0.017	0.00008
CPS Agg.	BC-CRSE, t(G-1)	0.218 (0.018)	0.017	0.0006
CPS Agg.	WC Bootstrap	0.24 (0.019)	0.017	0.003
CPS Agg.	FGLS	0.23 (0.018)	0.017	0.001
CPS Agg.	BC-FGLS	0.23 (0.018)	0.017	0.0002

Table 8: Rejection rates for tests of different methods for $\beta = 0.02$

Table of Contents

Robustness check: Varying Group Size

Now we vary the number of groups in the dataset to check the robustness of the methods to give a correctly sized test and detect real effects.

For every simulation we choose the group size number of states at random to form a sub sample out of the which, exactly half the states get treatment as before and the same treatment assignment procedure.

Robustness check: Varying Group Size

Method	G = 20	G = 10	G = 6
OLS (Assume iid)	0.287 (0.014)	0.229 (0.013)	0.071 (0.007)
CRSE,N(0,1)	0.073 (0.008)	0.085 (0.008)	0.117 (0.01)
BC-CRSE,t(G-1)	0.059 (0.007)	0.07 (0.008)	0.108 (0.01)
WC Bootstrap	0.052 (0.007)	0.06 (0.007)	0.075 (0.008)
FGLS	0.054 (0.007)	0.055 (0.007)	0.058 (0.007)
BC-FGLS	0.059 (0.007)	0.05 (0.006)	0.049 (0.006)

Table 9: Rejection rates for tests of nominal 5 percent size using different methods for different group using CPS Agg. data

Robustness check: Error Mis-specification

Here we conduct a Monte Carlo study using synthetically generated data to understand the rejection rate at $\beta = 0$ and $\beta = 0.02$ level.
We use a homogenous AR(1) data and the data generating process remains same as earlier. We also use a heterogenous AR(1) DGP where the first autocorrelation coefficient ρ is drawn from a uniform distribution $\mathcal{U}(0.2, 0.9)$ for each state.

$$y_{gt} = \alpha_g + \delta_t + \rho_g \cdot y_{gt-1} + e_{gt}$$

Thus here the homogenous autocorrelation structure is breaking down, which is the key assumption of FGLS process.
We also generate a homogenous MA(1) process with the equation:

$$y_{gt} = \alpha_g + \delta_t + \theta \cdot e_{gt-1} + e_{gt}$$

with $\theta = 0.5$

Robustness check: Mis-spefication of the Error Process

Method	Heterogenous AR(1)		Homogenous MA(1)	
	Type 1 Error	Power	Type 1 Error	Power
OLS (Assume iid)	0.362	0.227	0.137	0.142
CRSE, N(0,1)	0.055	0.046	0.046	0.06
BC-CRSE, t(G-1)	0.046	0.05	0.043	0‘.056
Res. Agg.	0.07	0.053	0.092	0.058
WC Bootstrap	0.04	0.08	0.04	0.06
FGLS	0.056	0.082	0.091	0.08
BC-FGLS	0.049	0.052	0.059	0.061

Table 11: Rejection rates for $\beta = 0$ and $\beta = 0.02$ respectively for hetrogenous AR(1) and homogenous MA(1) DGP for N = 50 and T = 20 for 1000 simulations